

Using Routinely Recorded Data in the UK to Assess Outcomes in a Randomised Controlled Trial

Powell, Graham; Bonnett, Laura; Smith, Catrin T.; Hughes, Dyfrig; Williamson, Paula; Marson, Anthony

Trials

DOI:

[10.1186/s13063-017-2135-9](https://doi.org/10.1186/s13063-017-2135-9)

Published: 23/08/2017

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):

Powell, G., Bonnett, L., Smith, C. T., Hughes, D., Williamson, P., & Marson, A. (2017). Using Routinely Recorded Data in the UK to Assess Outcomes in a Randomised Controlled Trial: The Trials of Access. *Trials*, 18, [389]. <https://doi.org/10.1186/s13063-017-2135-9>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Using Routinely Recorded Data in the UK to Assess Outcomes in a Randomised Controlled Trial: The Trials of Access

GA Powell; LJ Bonnett; CT Smith; DA Hughes; PR Williamson; AG Marson

Corresponding Author

Graham Powell, MRC HTMR Clinical Training Fellow
Department of Molecular and Clinical Pharmacology
Clinical Sciences Centre
Lower Lane
Fazakerley
Liverpool.
L9 7LJ.
Email: gpowell@liverpool.ac.uk
Telephone: 0151 529 5464

Co-authors

Laura J. Bonnett, NIHR Post-Doctoral Fellow, Department of Biostatistics, Waterhouse Building, Block F, 1-5 Brownlow Street, University of Liverpool, Liverpool, L69 3GL, U.K. ljbonnett@liverpool.ac.uk

Catrin Tudur Smith, Reader in Biostatistics, Department of Biostatistics, Waterhouse Building, Block F, 1-5 Brownlow Street, University of Liverpool, Liverpool, L69 3GL, U.K. cat1@liverpool.ac.uk

Dyfrig A. Hughes, Professor of Pharmacoeconomics, Centre for Health Economics & Medicines Evaluation, Institute of Medical and Social Care Research, College of Health and Behavioural Sciences, Bangor University, Ardudwy, Normal Site, Gwynedd, North Wales, LL57 2PZ, UK d.a.hughes@bangor.ac.uk

Paula R Williamson, Professor of Medical Statistics, Department of Biostatistics, Waterhouse Building, Block F, 1-5 Brownlow Street, University of Liverpool, Liverpool, L69 3GL, U.K. prw@liverpool.ac.uk

Anthony G. Marson, Professor of Neurology, Department of Molecular and Clinical Pharmacology, Clinical Sciences Centre, Lower Lane, Fazakerley, Liverpool. L9 7LJ, UK a.g.marson@liverpool.ac.uk

Abstract

Background

In the UK routinely recorded data may benefit prospective studies including randomised controlled trials (RCTs). In an on-going study, we aim to assess the feasibility of access and agreement of routinely recorded clinical and non-clinical data compared to data collected during a RCT using standard prospective methods. This paper will summarise available UK routinely recorded data sources and discuss our experience with the feasibility of accessing routinely recorded data for participants of a RCT before finally proposing recommendations for improving the access and implementation of routinely recorded data in RCTs.

Methods

Setting

The case study RCT is the Standard and New Antiepileptic Drugs II (SANAD II) Trial, a pragmatic, UK multicentre, phase IV RCT assessing the clinical and cost effectiveness of antiepileptic drug treatments for newly diagnosed epilepsy.

Participants

Ninety eight participants have provided written consent to permit the request of routinely recorded data.

Study Procedures

Routinely recorded clinical and non-clinical data were identified and data requested through formal applications from available data holders for the duration participants have been recruited into SANAD II. The feasibility of accessing routinely recorded data during a RCT is assessed and recommendations for improving access proposed.

Results

Secondary care clinical and socio-economic data is recorded on a national basis and can be accessed, although there are limitations in the application process. Primary care data is recorded by a number of organisations on a de-identified basis and access for specific individuals has not been feasible. Access to data recorded by non-clinical sources, including The Department for Work and Pensions and The Driving and Vehicle Licensing Agency, was not successful.

Conclusions

Recommendations discussed include further research to assess the attributes of routinely recorded data, an assessment of public perceptions and the development of strategies to collaboratively improve access to routinely recorded data for research.

Trial Registration

Standard and New Antiepileptic Drugs II (SANAD II (EudraCT No: 2012-001884-64, registered 05/09/2012; ISRCTN Number: 30294119, registered 03/07/2012))

Key words

Routine Data, Administrative Data, Feasibility, Data Collection

Background

There is a plethora of individual-level routinely recorded data in the UK. These data are recorded to fulfil specific, defined purposes and are regulated for security, confidentiality and disclosure by The Data Protection Act 1998 [1] and Freedom of Information Act 2000 [2]. Access to routinely recorded data for ‘secondary purposes’ such as clinical research is permitted providing there is demonstrable secondary benefit.

The potential for routinely recorded data to inform clinical research and Health Technology Assessment (HTA) has long been recognised [3]. Presently, there are a number of sources of routinely recorded primary and secondary care clinical data with regional or national coverage. However, limitations with accuracy of coding, confidentiality, ownership and data access have been previously identified as significant barriers to using routinely recorded data in research [4].

There are numerous examples of retrospective observational, record linkage population studies where routine sources have proved a valid and efficient method for providing data for clinical research [5]. In the context of prospective research such as randomised controlled trials (RCTs) routinely recorded data have been used to inform judgements about the feasibility of sample size and recruitment targets [6] and measuring participant outcomes [3, 7]. Pragmatic cluster RCTs have been coordinated through routine data sources including patient recruitment, randomisation, administration of intervention and trial assessments, such as through the Clinical Practice Research Datalink (CPRD) [8]. The majority of RCTs incur costs as clinicians assess participants, record outcomes and complete Case Report Forms - hence using routinely recorded data may provide an efficient alternative method for data collection in addition to reducing the burden on participants. Furthermore, data from non-clinical routine sources may inform outcomes beyond the standard RCT assessments of clinical efficacy and effectiveness. For example, cost data (such as use of healthcare resources) and socio-economic data (such as employment and means-tested benefits data) may inform health economic analyses and the assessment of the broader societal impact of healthcare interventions.

The potential benefits of using routinely recorded data in clinical research have resulted in a political drive to increase implementation, detailed in *The Plan for Growth* [9] and *The NHS constitution* [10], where research is presented as a core activity making the link explicit between the provision of NHS services and research. Consequently, initiatives such as the Administrative Data Research Network [11] have been established to provide a method of access to individual-level data, linking clinical and non-clinical sources of routinely recorded data.

The objective of this paper is to review relevant sources of routinely recorded data for England, Scotland and Wales and discuss our experience with the feasibility of accessing individual-level data for a subgroup of participants enrolled into a RCT before finally proposing recommendations for improving the access and implementation of routinely collected data in RCTs. This is an on-going study and in a future publication we aim to assess the agreement of routinely recorded data compared to paired data collected in a RCT using standard prospective methods.

Methods

The case study RCT is the Standard and New Antiepileptic Drugs (SANAD) II Trial. SANAD II is a pragmatic, UK multicentre, phase IV RCT funded by the National Institute for Health Research (NIHR) HTA programme, assessing the clinical and cost effectiveness of a number of antiepileptic drugs as first line treatments for newly diagnosed epilepsy. Data for clinical outcomes including seizure freedom and adverse events are recorded on Case Report Forms by the treating clinical team during outpatient appointments. Data to inform cost effectiveness analyses including healthcare resource use and quality of life are recorded through participant completion of questionnaires. SANAD II is currently recruiting and is expected to report in 2019.

Following research ethics and governance approvals, 470 participants enrolled in SANAD II were invited to provide written consent to permit the request of routinely recorded data for the duration of their participation in SANAD II. Ninety eight (20.9%) participants provided consent and were included in the study. Relevant sources of routinely recorded data were identified and detailed scoping discussions ensued. Subsequently where accessible, routinely recorded data for participants recruited into SANAD II were requested through formal applications.

- The routinely recorded data sources included in this study are as follows: Clinical Routine Data Sources: Secondary Care:
 - o The Health and Social Care Information Centre (HSCIC)
 - o The NHS Wales Informatics Service (NWIS)
 - o The NHS National Services Scotland; Information Services Division (ISD)
- Clinical Routine Data Sources: Primary Care:
 - o The Clinical Practice Research Datalink (CPRD)
 - o ResearchOne
 - o QResearch
 - o The Health Improvement Network (THIN) Database
 - o North West eHealth (NWEH)
- Non-Clinical Routine Data Sources:
 - o The Office for National Statistics (ONS)
 - o HM Revenue and Customs (HMRC)
 - o The Department for Work and Pensions (DWP)
 - o The Driver and Vehicle Licensing Authority (DVLA)
- 'Linked' Routine Data Sources:
 - o The Secure Anonymised Information Linkage (SAIL) Databank
 - o The Administrative Data Research Network (ADRN)

In a future publication, the agreement between routinely recorded data and data collected using standard prospective methods will be assessed for baseline variables such as gender, age and date of first seizure and for outcome measures relevant to SANAD II such as time to 12 month remission from seizures. To assess agreement between paired continuous data, Bland Altman methods will be employed. Acceptable clinical limits of agreement for each variable or SANAD II outcome will be specified *a priori* and compared to the 95% confidence limits of agreement. To assess agreement between paired, nominal categorical datasets, cross tabulations will be constructed followed by calculation of Cohen's Kappa.

Results

Clinical Routine Data Sources: Secondary Care

Electronic medical records of patients' use of secondary care services in the UK are routinely managed on a national basis. A number of public service organisations provide national information, data and IT systems for commissioners, analysts and clinicians in health and social care. Data is recorded to inform patient care, provide the data for remuneration for Hospital Trusts and is subsequently used to monitor and improve clinical services through clinical research. *Table 1* summarises the data sources where access to individual-level data is possible.

Table 1: Example sources of Routinely Recorded Secondary Care Data

Clinical Routine Data Sources: Primary Care

Electronic medical records of patients' use of primary care services in the UK are recorded routinely by the General Practitioner to inform patient care and remuneration, but are not currently available for clinical research on a national basis. A number of organisations represent collaborations between governmental bodies or academic institutions and providers of Primary Care IT systems. Access on a regional basis is possible through a number of data sources summarised in *Table 2*.

Table 2: Example sources of Routinely Recorded Primary Care Data

Non-Clinical Routine Data Sources

Non-clinical, individual-level data are routinely recorded by a number of UK governmental departments for a variety of indications. Selected organisations record data that would be informative to prospective clinical research in epilepsy and other diseases, summarised in *Table 3*.

Table 3: Example sources of Routinely Recorded Non-Clinical Data

‘Linked’ Routine Data Sources

In order to provide a ‘complete’ dataset of the information required to meet research objectives, data from a number of organisations may need to be accessed. This is typically accomplished by linking data sources using identifiers such as patients’ name, date of birth, National Insurance Number or NHS Number. In response to the growing recognition of the potential of routinely recorded data, initiatives have been established to assist with the provision of linked, de-identified, aggregate data between data sources:

- **The Secure Anonymised Information Linkage (SAIL) Databank** is an initiative developed by Swansea University and funded by the Welsh Government. SAIL provides a method of access to individual-level, routinely-recorded, de-identified electronic data for patients across Wales to support research [36]. Access to clinical datasets provided by NWIS is complemented with numerous non-clinical administrative datasets including births, deaths and demographic data. Following the scoping process a formal application is submitted to the *Information Governance Review Panel* before access to data is granted. SAIL data has been accessed to measure clinical outcomes in retrospective research [37].
- **The Administrative Data Research Network (ADRN)** is a UK-wide partnership between universities, government departments, national statistics authorities, funders and researchers, funded by the *Economic and Social Research Council*. ADRN provides a method of access to a number of non-clinical administrative routine datasets including employment, socioeconomic, crime and education data [11] in addition to clinical datasets detailed previously such as those recorded by the HSCIC. Following development of a project proposal a formal application is reviewed by the *Approvals Panel* before access to data is granted.

Challenges and Feasibility of Access

We have requested access to routinely recorded data for individuals enrolled in the SANAD II RCT, resident in England and Wales who have provided written consent. There were insufficient participants meeting the eligibility criteria resident in Scotland. Data sources were identified and scoping discussions informed the initial assessment of feasibility. Data sources were deemed feasible if individual-level data could be provided for specified individuals providing consent. Resources required including cost and researcher time were also factors important in the assessment of feasibility. Including the preparation, research ethics and governance approval and submission of the applications for data access, significant researcher time and a period of 18 months were required. The feasibility, timeline and key milestones involved for each data source are summarised in *Table 4*.

Clinical Routine Data Sources

Routinely recorded secondary care data can be requested on an individual-level, identifiable basis for patients in England and Wales through HSCIC and NWIS, accessed through SAIL and in our experience this process is feasible as part of a RCT, yet there are notable limitations. In England, the HSCIC has set a target time to data access of 60 working days following submission for a complex application, involving bespoke data linkage from multiple datasets. From the date of submission of the Data Access Request Service online application we have been granted access to the data within this timeframe. However, this positive experience following submission of the application is countered by limitations in the pre-application process. Acknowledging the significant update to online application and approval procedures that occurred during this period, there remains a considerable period of time required in the development of the application. The nature of the request for identifiable data necessitated participant consent as the valid legal basis. The HSCIC require ethical and governance approval to be in place prior to DARS review and to prevent future amendments and delays, it was rational to ensure the consent materials had been reviewed by the HSCIC Information Governance team, prior to submitting the documents for ethical and governance approval. HSCIC provide written guidance regarding the consent materials and advise documents should be reviewed. However, in our experience there is no formalised process for providing this review. Following significant correspondence the consent materials were reviewed by the Data Access and Information Sharing Team. However, this feedback was provided following a formal submission and review by the Data Access Request Service. Formalising the process for the review of consent materials would likely improve the time and resource efficiency for both HSCIC and the researcher.

For participants in Wales, we have requested secondary care data and for a proportion of participants primary care data through SAIL Databank. SAIL provided a streamlined pre-application service, including engaging in multiple discussions and completion of a scoping document, outlining the study methods and costs involved. Consent materials were also promptly reviewed by a member of the Information Governance team.

Common to both sources of secondary care routinely recorded data; there are stringent information governance requirements that must be in place prior to application. These include information security measures and assessments, specific inclusion regarding the ‘processing of healthcare data for the subjects of research’ in the institutional Data Protection Act registration and in the case of HSCIC, an institutional Data Sharing Framework Contract. Adequate guidance is provided by the data sources and if not addressed by the researcher, may cause delay. Furthermore, there is a time lag of approximately 3-6 months before data becomes available within each data source. This delay

potentially limits the utility of such sources in prospective clinical research, such as drug trials where prompt reporting is clinically important and a regulatory requirement.

Routinely recorded primary care data for specific participants in England is less accessible. The majority of providers of primary care data such as ResearchOne and QResearch provide data on a de-identified basis with no facility to re-identify individuals. Therefore, where specific participants need to be identified, as for RCTs such as SANAD II, these sources are not applicable. Following our correspondence, CPRD confirmed it may be possible to retrieve identifiable individual-level data, linked to HSCIC data in the future, but the required approvals were not in place and the timescale to resolution was unclear. Furthermore, such primary care sources provide data for only a proportion of the population and can be expensive.

NorthWest eHealth employs an alternative methodology whereby primary care data is extracted directly from the GP through a third party. This process requires participant and GP consent and installation of the required software but is an effective data extraction method [27]. NWEH offers a number of primary care research tools for the wider research community but does not currently routinely provide a bespoke primary care data extraction service for research.

Non-Clinical Routine Data Sources

Aggregate economic and societal statistics, provided by Lower Layer Super Output Area (LSOA) can be accessed through the ONS and are in the public domain. Such data may have additional benefits to the analyses of health and socio-economic outcomes in RCTs. Individual-level economic data from sources such as DWP and HMRC would likely be informative to prospective clinical research as such data are often poorly or incompletely recorded using standard methods [38]. However, relevant to this study, there is no previous evidence of access to DWP or HMRC individual-level or aggregate data for clinical research.

During scoping discussions with DWP and HMRC, we were directed to the ADRN but this network has not been successful in negotiating data access.

Finally, the outcomes of selected clinical studies may be measured using DVLA data. However, the DVLA declined the request for access, citing insufficient internal resources to process the request and more stringent data protection requirements than those employed in the NHS or academic institutions, without providing explicit details regarding these requirements.

Table 4: Summary of Key Application Milestones

Discussion

Routinely recorded data are valid for use in retrospective clinical research [3, 4] and have the potential to be used in prospective research including measuring the outcomes of RCTs [7] and providing additional benefits such as a method to address missing RCT data. Limitations, specifically with respect to accuracy and access have been recognised for some time. Academic, political [9] and health service [10] interest in UK sources of routinely recorded data has resulted in expansion and improvements, notably in the access to linked datasets. However our experience with accessing individual-level data for specific participants providing written consent to inform the outcomes of a RCT, highlight persisting limitations.

Clinical routine data sources are numerous and there is comprehensive national coverage of secondary care data. In our experience, accessing individual level data is feasible. However, inefficiencies in the application processes persist, particularly during the informal ‘pre-application’ phase. The notable limitation encountered was obtaining feedback on the Patient Information Sheet and Consent Form prior to ethical and governance review. Formalising an explicit review process for consent materials would improve the efficiency for both the data holders and the research team.

Access to routinely recorded individual - level primary care data has not been feasible. Each primary care data source has limited geographical coverage often based on GP IT system, usually process de-identified data and may incur significant expense. The inception of the HSCIC *General Practice Extraction Service* which records primary care data nationally for England, represents the most optimistic national source, however access is currently restricted to Department of Health initiatives, such as research involving screening procedures [39].

The access to non-clinical data sources for clinical research has not been possible. The ADRN has been established to act on behalf of the researcher in negotiating access to de-identified, linked routinely recorded data from a number of organisations and the study proposal was promptly directed to the ADRN. However, the decision whether to release data remains with the data holder. Ideologically, the next step would be the storage of de-identified linked data from participating organisations in a single repository, similar to those established for RCT data [40]. This would create a single point of access and remove the burden for each organisation to consider each study individually. This would however require significant information governance and security barriers to be cleared and in light of recent developments within the research climate, individual consent. Including patients as stakeholders in the development of such data sources is essential [41].

Although there are examples of pragmatic RCTs being coordinated through routine data sources [8], there are likely to be limitations when accessing routinely recorded data to measure the outcomes of RCTs. Quality assurance is unclear and the level of agreement of routinely recorded data with data recorded through standard RCT methods remains uncertain, particularly when measuring clinical outcomes. The time delay before routinely recorded data becomes available may have implications for RCTs where prompt reporting is both clinically important and a regulatory requirement. Furthermore the pre-application and application process may introduce further delays. This will have implications for RCTs relying on routinely recorded data. The cost efficiency of accessing routinely recorded data, compared to standard methods is unclear. Further research is required to assess the agreement, additional benefits and cost efficiency of routinely recorded data compared to data collected through standard RCT methods; it may be in the additional benefits such as addressing missing RCT data where routinely recorded data is most useful.

Conclusions

The failure of access to routinely recorded data for a purpose such as this study with clear secondary benefit to clinical research methodology seems inappropriate when the ‘public purse’ funds the research, the researcher and the public body holding the data. Perhaps a significant cause or contributor to the current limitations is the Care.Data initiative in 2014. The proposal to extract primary care records from all patients was opposed publicly by a number of groups and for example, resulted in an internal inquiry within the HSCIC. Data applications were suspended during this period and our current experience may be explained by the concurrent revision of the HSCIC application and approval procedures. However, in the medium term, of more concern is the harm in public perception that resulted. Currently, more than 1.2 million individuals in the UK have submitted a ‘Type 2 objection’, meaning that their data will not be shared for purposes other than direct care [42]. Although the application procedures may improve and in time we may be able to access data more efficiently, the loss of 2.2% of the population will have implications for the routinely recorded data that will then be made available for research. Involving patients as important stakeholders and re-gaining their trust will be an essential factor in realising the individual and population healthcare benefits of routinely recorded data [43].

Recommendations

We propose recommendations to improve access and implementation of routinely recorded data during a RCT, summarised in *Table 5*.

Table 5: Recommendations to Improve Access to Routinely Recorded Data for Research

List of Abbreviations

ADRN (Administrative Data Research Network)

CPRD (Clinical Practice Research Datalink)

DVLA (Driver and Vehicle Licensing Agency)

DWP (Department for Work and Pensions)

HMRC (HM Revenue and Customs)

HSCIC (The Health and Social Care Information Centre)

ISD (Information Services Division)

NHS (National Health Service)

NWEH (North West EHealth)

NWIS (NHS Wales Information Centre)

ONS (The Office for National Statistics)

RCT (Randomised Controlled Trial)

SANAD II (The Standard and New Antiepileptic Drugs Trial II)

SAIL (The Secure Anonymised Information Linkage Databank)

THIN (The Health Information Network)

Declarations

Ethics Approval and Consent to Participate

To request identifiable data from routine data sources, written participant consent was required. The North of Scotland Research Ethics Service (16/NS/0007) and Health Research Authority (IRAS 189002) approved the study.

Consent for Publication

Not Applicable

Availability of Data and Materials

There is no dataset available for this study and therefore data sharing is not applicable.

Competing Interests

All authors declare that (1) GAP, LJB, CTS, DAH, PRW and AGM do not have support from any company for the submitted work; (2) GP, LJB, CTS, DAH, PRW and AGM have no financial relationship with any company that might have an interest in the submitted work in the previous three years; and (3) GP, LJB, CTS, DAH, PRW and AGM have no other relationships or activities that could appear to have influenced the submitted work.

Funding

This report is independent research arising from a Clinical Training Fellowship (GA Powell) awarded by the Medical Research Council Hubs for Trials Methodology Research (Reference: P16-2014-GPC). The views expressed are those of the authors and not necessarily those of the NHS, the Medical Research Council, or the routine data sources detailed in the report.

Authors Contributions

Graham Powell (GAP) performed the background research, prepared the protocol, liaised with the routine data sources, prepared and submitted the routine data applications and drafted and redrafted the manuscript. Laura Bonnett (LJB), Catrin Tudur Smith (CTS), Dyfrig Hughes (DAH), Paula Williamson (PRW) and Anthony Marson (AGM) provided input into the development of the protocol and drafted and redrafted the manuscript. AGM is the guarantor for the report. The University of Liverpool was the sponsor for the study. All authors read and approved the final manuscript.

Acknowledgements

Not Applicable

References

1. UKGovernment. *The Data Protection Act*. 2016 Accessed March 2016 [cited 2016 03/01/2016]; Available from: (<https://www.gov.uk/data-protection/the-data-protection-act>)
2. UKGovernment. *The Freedom of Information Act*. 2016 [cited 2016 04/03/2016]; Available from: (<http://www.legislation.gov.uk/ukpga/2000/36/contents>)
3. Lewsey, J.D., et al., *Using routine data to complement and enhance the results of randomised controlled trials*. Health Technology Assessment, 2000. **4**(22): p. i+iii-iv+1-45.
4. Raftery, J., P. Roderick, and A. Stevens, *Potential use of routine databases in health technology assessment*. Health technology assessment (Winchester, England), 2005. **9**(20): p. 1-92, iii-iv.
5. Clarson, L.E., et al., *Increased risk of vascular disease associated with gout: A retrospective, matched cohort study in the UK Clinical Practice Research Datalink*. Annals of the Rheumatic Diseases, 2015. **74**(4): p. 642-647.
6. McGregor, J., et al., *The Health Informatics Trial Enhancement Project (HITE): Using routinely collected primary care data to identify potential participants for a depression trial*. Trials [Electronic Resource], 2010. **11**: p. 39.
7. Williams, J.G., et al., *Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment*. Health technology assessment (Winchester, England), 2003. **7**(26): p. iii, v-x, 1-117.
8. Gulliford, M.C., et al., *Cluster randomized trials utilizing primary care electronic health records: Methodological issues in design, conduct, and analysis (eCRT Study)*. Trials, 2014. **15**(1).
9. *The Plan for Growth*. 2011.
10. *The NHS Constitution for England*. 2013.
11. ADRN. *The Administrative Data Research Network*. 2016 [cited 2016 16/06/2016]; Available from: <http://adrn.ac.uk>.
12. HSCIC. *The Health and Social Care Information Centre*. 2016 [cited 2016 16th July]; Available from: <http://www.hscic.gov.uk>.
13. Bouras, G., et al., *Linked hospital and primary care database analysis of the incidence and impact of psychiatric morbidity following gastrointestinal cancer surgery in England*. Annals of Surgery, 2016. **264**(1): p. 93-99.
14. Turner, E.L., et al., *Design and preliminary recruitment results of the Cluster randomised triAl of PSA testing for Prostate cancer (CAP)*. British Journal of Cancer, 2014. **110**(12): p. 2829-36.
15. NWIS. *NHS Wales Informatics Service*. 2016 [cited 2016 04/05/2016]; Available from: <http://www.wales.nhs.uk/nwis/page/52490>.
16. Ismail, S.I. and B. Puyk, *The rise of obstetric anal sphincter injuries (OASIS): 11-year trend analysis using Patient Episode Database for Wales (PEDW) data*. Journal of Obstetrics & Gynaecology, 2014. **34**(6): p. 495-8.
17. ISDSScotland. *The Information Services Division* 2016 [cited 2016 05/05/2016]; Available from: <http://www.isdscotland.org/Products-and-Services/index.asp>.
18. Ahmed, A., et al., *Upper gastrointestinal bleeding in Scotland 2000-2010: Improved outcomes but a significant weekend effect*. World Journal of Gastroenterology, 2015. **21**(38): p. 10890-7.
19. CPRD. *The Clinical Practice Research Datalink*. 2016 [cited 2016 04/05/2016]; Available from: <http://www.cprd.com/intro.asp>.
20. TPP. *ResearchOne*. 2016 [cited 2016 04/05/2016]; Available from: <http://www.tpp-uk.com/products/systmone>.
21. Herrett, E., et al., *Text messaging reminders for influenza vaccine in primary care: protocol for a cluster randomised controlled trial (TXT4FLUJAB)*. BMJ Open, 2014. **4**(5): p. e004633.
22. QResearch. *QResearch*. 2016 [cited 2016 06/05/2016]; Available from: <http://www.qresearch.org/SitePages/Home.aspx>.

23. Hill, T., et al., *Antidepressant use and risk of epilepsy and seizures in people aged 20 to 64 years: cohort study using a primary care database*. BMC Psychiatry, 2015. **15**: p. 315.
24. THIN. *The Health Improvement Network*. 2016 [cited 2016 05/04/2016]; Available from: <http://www.thin-uk.net>.
25. González-Pérez, A., et al., *Incidence and Predictors of Hemorrhagic Stroke in Users of Low-Dose Acetylsalicylic Acid*. Journal of Stroke and Cerebrovascular Diseases, 2015. **24**(10): p. 2321-2328.
26. NWEH. *North West eHealth*. 2016 [cited 2016 05/05/2016]; Available from: <http://nweh.org.uk>.
27. New, J.P., et al., *Obtaining real-world evidence: The Salford Lung Study*. Thorax, 2014. **69**(12): p. 1152-1154.
28. ApolloMedical. *Apollo Data Extraction*. 2016 [cited 2016 06/04/2016]; Available from: <http://www.apollo-medical.com/>.
29. GraphnetHealth. *Graphnet*. 2016 [cited 2016 06/04/2016]; Available from: <http://www.graphnethealth.com/what-we-do/overview/what-we-do>.
30. ONS. *The Office for National Statistics*. 2016 [cited 2016 04/04/2016]; Available from: <https://www.ons.gov.uk>.
31. ONS. *Official Labour Market Statistics*. 2016 [cited 2016 04/04/2016]; Available from: <http://www.nomisweb.co.uk>.
32. OCSI. *Data 4 Neighbourhoods and Regeneration 2016* [cited 2016 07/04/2016]; Available from: <http://www.data4nr.net/introduction>.
33. UKGovernment. *HM Revenue and Customs*. 2016 [cited 2016 04/04/2016]; Available from: <https://www.gov.uk/government/organisations/hm-revenue-customs>.
34. UKGovernment. *The Department for Work and Pensions*. 2016 [cited 2016 01/05/2016]; Available from: <https://www.gov.uk/government/organisations/department-for-work-pensions>.
35. UKGovernment. *Driver and Vehicle Licensing Agency*. 2016 [cited 2016 02/05/2016]; Available from: <https://www.gov.uk/government/organisations/driver-and-vehicle-licensing-agency>.
36. SAIL. *The Secure Anonymised Information Linkage Databank*. 2016 [cited 2016 01/06/2016]; Available from: <http://www.saildatabank.com>.
37. Sayers, A., et al., *Evidence for a persistent, major excess in all cause admissions to hospital in children with type-1 diabetes: results from a large Welsh national matched community cohort study*. BMJ Open, 2015. **5**(4): p. e005644.
38. (US), N.-R.-C., *The Prevention and Treatment of Missing Data in Clinical Trials*, in *National Academies Press (US)*. 2010.
39. HSCIC. *General Practice Extraction Service*. 2016 [cited 2016 05/04/2016]; Available from: <http://www.hscic.gov.uk/gpes>.
40. Clinical-Study-Data-Request. *Clinical Study Data Request*. 2016 [cited 2016 05/04/2016]; Available from: <https://www.clinicalstudydatarequest.com>.
41. Eugene C Nelson, M.D.-W., Paul B Batalden, Karen, A.D.V.C. Homa, Tamara S Morgan, Elena, and E.S.F. Eftimovska, John Ovretveit, Wade Harrison, Cristin Lind Staffan Lindblad, *Patient focused registries can improve health, care and science*. BMJ, 2016. **354**.
42. HSCIC. *Press Release: Patient Opt Out*. 2016 [cited 2016 01/05/2016]; Available from: <http://www.hscic.gov.uk/catalogue/PUB20527>.
43. Tjeerd-Pieter Van Staa, I.B., Ben Goldacre, Liam Smeeth, *Big health data: the need to earn public trust after past management*. BMJ, 2016. **354**: p. 95-97.
44. Clarke, M., *Standardising outcomes for clinical trials and systematic reviews*. Trials, 2007. **8**.

List of Tables

Table 1: Example sources of Routinely Recorded Secondary Care Data

Table 2: Example sources of Routinely Recorded Primary Care Data

Table 3: Example sources of Routinely Recorded Non-Clinical Data

Table 4: Summary of Key Application Milestones

Table 5: Recommendations to Improve Access to Routinely Recorded Data for Research

